Ethics of "Taste, Ties and Time"

Linzhuo Li

1. Summary

In this paper, I analyze how the Facebook study conducted failed to follow Salganik's four principles of research ethics and how one might improve the research design in the future research to better protect the privacy of the data subjects and reduce potential harms on them. I also discussed several aspects that should be considered if I am doing a research with this data set.

2. The Four Principles of ethical research

In "Bit-by-bit", Salganik gives four principles that are related with research ethics: Respect for persons, Beneficence, Justice, Respect for law and public interest. Reviewing each of these principles one by one makes it clear that the Facebook study "Taste, Ties, and Time"(T3), although being aware of some aspects of research ethics, still violated quite a few points related to the four principles and would need to be redesigned if one is to repeat a similar study.

2.1 Respect for persons:

As the first principle, "respect for persons" requires research designers to draw more attention on the subjects of the data. While it includes things such as treating subjects as autonomous individuals and protect their autonomy, a key ethical point is to make sure the research is gone with the consent of the data subjects. In "T3", we can see that this principle is violated. First, the researchers didn't get the consent of the students in study. They scraped the Facebook profiles and merged it with students' information in the college, including their residential and academic information. Although their research was approved by IRB at Harvard, it still needs to respect students' autonomy in a better way. For example, as suggested in the book, the researchers could have gotten some form of consent from the students, even only a percentage of them, they can send emails for example to them to get their feedback on whether or not they are feeling comfortable when their information offline is merged with their traces online. They should also notice that getting permission from Facebook—the service provider and the University—-part of

the data holder—- doesn't equal to getting permissions from data subjects. The researchers not only used the merged data to conduct their own research, but also shared it online with other researchers, again without getting the consent of data subjects. This can be potentially risky especially when others used this data for non-academic purposes. This potential danger is not considered by the researchers.

2.2 Beneficence

The two fundamental parts of "Beneficence" is a) do not harm and b) maximize possible benefits and minimize possible harms. As far as I can tell, the "T3" study could have done better to prevent many of the potential harms it brought to subjects. For instance, although the researchers claimed that "all identifying information was deleted or encoded", the University's name and class group were still every quickly recognized by others. This makes it possible to actually trace and identify each individual of data subjects, which would put these individual's privacy and their normal social activity at risk. To make the source of data more in protection for example, instead of saying "the freshman class of 2009 at a diverse private college in the Northeast U.S", they could have not shown the location, the cohort information and the characteristics of the college, making the data source more disclosed.

Moreover, the researchers should be more aware of the potential risks that are related with the race, ethnicity, courses they take and their political views information, as these information can be considered as private by subjects or they may not want to let it go public. Furthermore, following the principle of doing less harm, they should have a more consistent thought on what forms of data might lead to disclosure of personal information and how they can avoid post-hoc re-identification, especially by those within the subject group.

2.3 Justice

As noted in the textbook, the benefit and cost should be balanced among interest groups related with the research. While it is debatable whether college students under study in T3 are "vulnerable" and whether or not they can benefit from this study, it is certain that they should not be harmed. It is also certain that students should not be the only group that bears the burden of

this study while research groups and other organizations(say, advertising companies) benefit. Regarding this, we may think whether and how these students can be benefited from this research.

2.4 Respect of laws and Public interest

Similar to beneficence, but still different, respect of laws and public interest focus not just on subjects but also on public and laws. In my interpretation, this means researchers should be more aware of the legal and social consequences of their research. They need to be careful at all stages of their research whether their methods are illegal or morally unacceptable. In the case of T3, the researchers could have consulted a third-party ethical review before they made their dataset public.

3. Whether or not I would use the data

Despite that the comprehensiveness of the data can generate great findings and contribute potentially a lot to sociological theories, I personally would not take this dataset in my own research. The main concern is that unlike some other datasets, the personal information of the subjects of study on Facebook are real and can be mapped/merged with their information in the real world. To me, the distance between digital world and real world is one of the important reasons why people in digital world such as Facebook feels free to be active and express their feelings. Sometimes they are not aware of the fact that their online behaviors can be traced and linked to their offline life. They more or less separate their activities and the "role" they play online from offline. As researchers, we need to think from their perspectives and try not do harm to their offline life. The data in its own cannot guarantee that. Even after some coding and encryption, it is not 100% guaranteed that subjects are unidentifiable. For example, as one has argued, "in digital age, network itself is a finger print". Researchers of T3 are over optimistic of the confidentiality of their data, and the unexpected criticism after their research really teaches a lesson to all of us.